2019年6月5日　統計数理研究所　オープンハウス

２変量分布の上側と下側の裾確率を比較するための尺度

加藤　昇吾　数理・推論研究系　准教授

はじめに

背景

確率分布をデータにあてはめるとき，裾の部分のあてはまりが悪いと，誤った解析結果を導くことがある．

そこで，裾の重さを調節できる分布が，今までに数多く提案されてきた．

特に近年は，上側と下側の裾が対称ではない “非対称な裾” を持つ分布の研究が盛んである．（例：skew-t 分布，クレイトンコピュラ，グンベルコピュラ）

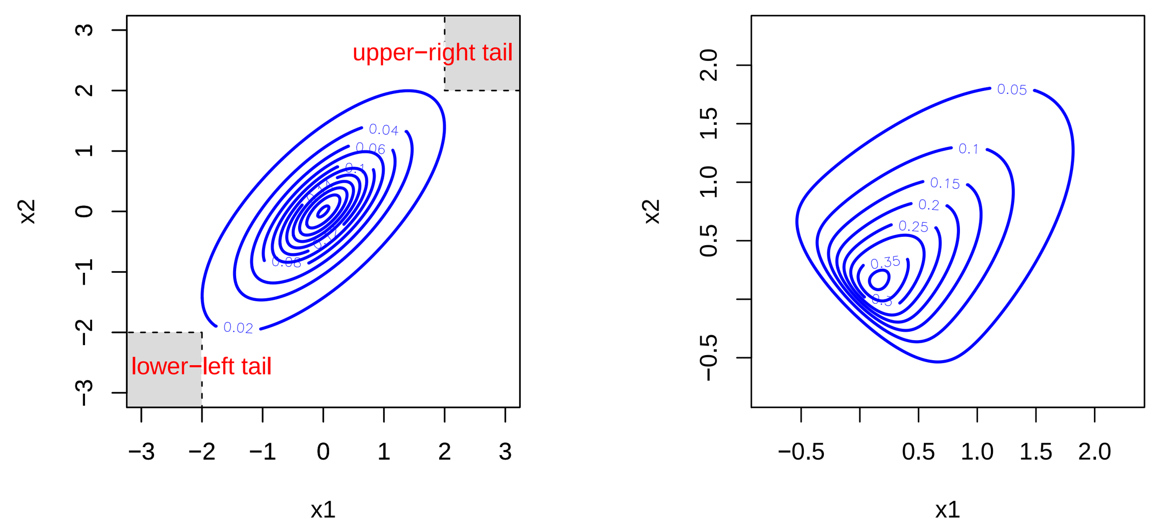


図１．自由度1の t 分布（左）と skew-t 分布（右）の確率密度関数の等高線図．

研究の目的

データに分布を当てはめるとき，対称な裾・非対称な裾のどちらを持つ分布を当てはめれば良いかを見極めることは難しい．

そこで本報告では，２変量分布の上側と下側の裾確率を比較するための尺度を提案する．

なお本研究は，吉羽要直特任教授（首都大学東京），江口真透教授（統計数理研究所）との共同研究である．

裾確率を比較するための尺度

定義

(X_1, X_2): ２変量確率ベクトル，
F_j: X_j の（連続型）分布関数 $(j = 1, 2)$，
とする．このとき，(X_1, X_2) の分布の左下側と右上側の裾確率を比較するための尺度を，次のように定義する．

$$\alpha(u) = \log\left(\frac{P(F_1(X_1) > 1-u, F_2(X_2) > 1-u)}{P(F_1(X_1) \le u, F_2(X_2) \le u)}\right), \quad 0 < u \le 0.5.$$

$\alpha(u)$ の性質

性質１

C を (X_1, X_2) のコピュラとする，つまり，

$$C(u_1, u_2) = P(F_1(X_1) \le u_1, F_2(X_2) \le u_2),$$

である．このとき，$\alpha(u)$ は次のようにあらわすことができる．

$$\alpha(u) = \log\left(\frac{2u - 1 + C(1-u, 1-u)}{C(u, u)}\right).$$

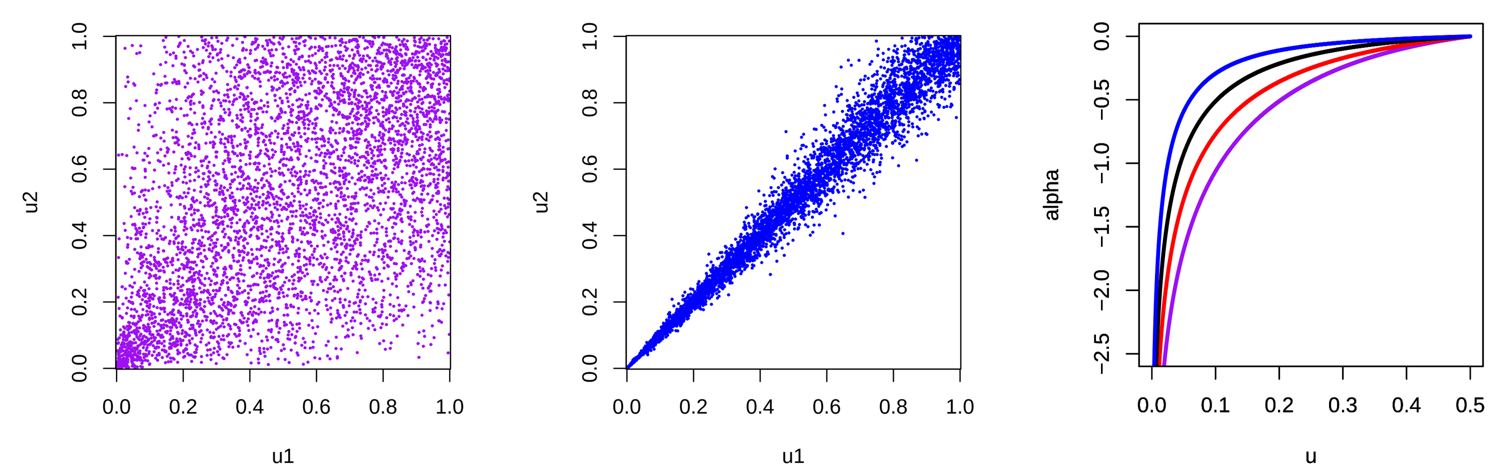
- （２変量）コピュラとは，それぞれの周辺分布が $[0, 1]$ 上の一様分布となる２変量分布関数のことをいう．
- $C(u_1, u_2)$ の生存関数を $\bar{C}(u_1, u_2) = 1 - u_1 - u_2 + C(u_1, u_2)$ とあらわせば，$\alpha(u)$ はより簡潔に以下のように表現できる．

$$\alpha(u) = \log\left(\frac{\bar{C}(1-u, 1-u)}{C(u, u)}\right).$$

性質２

$\alpha(u)$ について以下が成立する：

(i) $\alpha(u) = 0 \iff C(u, u) = \bar{C}(1-u, 1-u)$,
(ii) $\alpha_C(u) = -\alpha_{\bar{C}}(u)$,
(iii) $\alpha(0.5) = 0$,
(iv) C に関するある条件の下で，$\lim_{u \downarrow 0} \alpha(u) = \log(\lambda_U / \lambda_L)$ が成り立つ．ここに，λ_U（λ_L）は C の上（下）裾依存係数である．



$$C(u_1, u_2) = \max\{u_1^{-\theta} + u_2^{-\theta} - 1, 0\}^{-1/\theta}, \quad \theta \in [-1, \infty) \setminus \{0\}.$$

図２．クレイトンコピュラから発生した擬似乱数のプロット.（左）$\theta = 1$,（中）$\theta = 20$.（右）クレイトンコンピュラの $\alpha(u)$ のプロット（$\theta = 1$, $\theta = 5$, $\theta = 10$, $\theta = 20$）.

$\alpha(u)$ の推定量

$\{(U_{1i}, U_{2i})\}_{i=1}^n \sim$ i.i.d. $C(u_1, u_2)$ とする．このとき，$\alpha(u)$ の推定量を

$$\hat{\alpha}(u) = \log\left(\frac{T_U}{T_L}\right),$$

と定義する．ここに，

$$T_U = \frac{1}{n}\sum_{i=1}^n \mathbf{1}(U_{1i} \ge 1-u, U_{2i} \ge 1-u), \quad T_L = \frac{1}{n}\sum_{i=1}^n \mathbf{1}(U_{1i} \le u, U_{2i} \le u).$$

$\hat{\alpha}(u)$ の性質

$\mathbb{A}_n(u)$ を以下で定義する．

$$\mathbb{A}_n(u) = \sqrt{n}\,\{\hat{\alpha}(u) - \alpha(u)\}, \quad 0 < u \le 0.5.$$

このとき，$n \to \infty$ とすると，$\{\mathbb{A}_n(u) \,|\, 0 < u \le 0.5\}$ は平均関数 0，共分散関数 $\sigma^2_{u \vee v}$ のガウス過程に弱収束する．ただし，$u \vee v = \max(u, v)$,

$$\sigma_w^2 = \frac{C(w, w) + \overline{C}(1-w, 1-w)}{C(w, w) \cdot \overline{C}(1-w, 1-w)}.$$

上記の結果を応用すると，$\alpha(u)$ の漸近信頼区間を得ることができる．

株価データへの応用

- 2005年4月1日から2015年3月31日まで記録された日経225とS&P500のdaily returnを考える $(n = 2367)$．
- 日経225とS&P500のdaily returnにGARCH(1,1)モデルを当てはめたときの標準化された残差 $\{(x_i, y_i)\}$ について，$\alpha(u)$ を推定する．

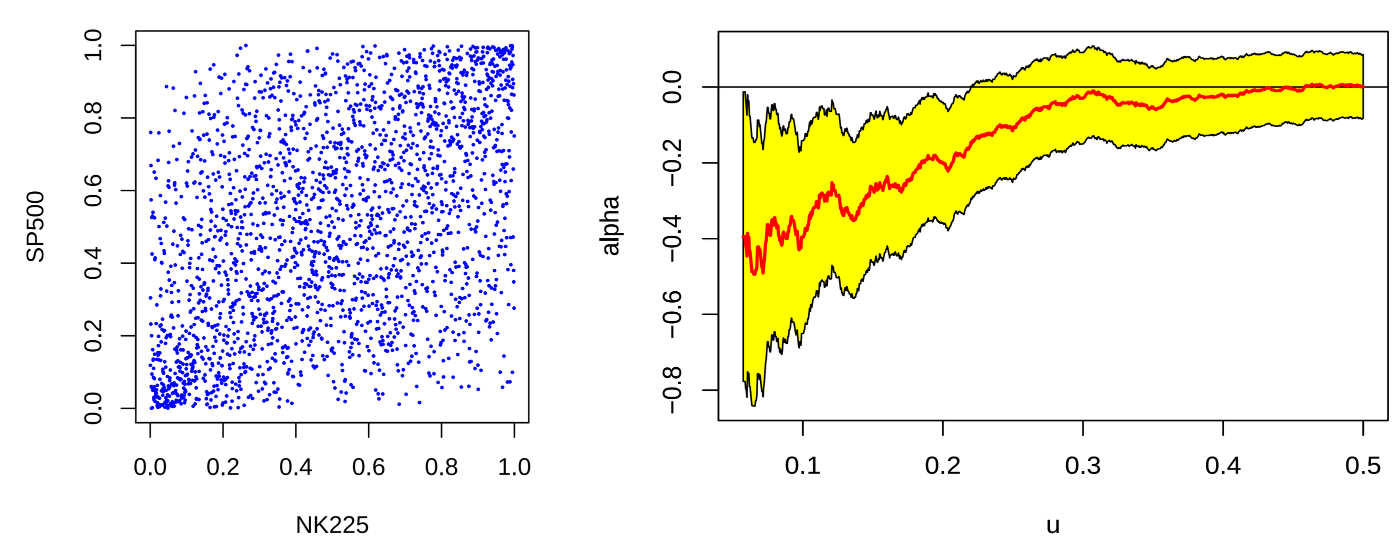


図３.(左) $\{\hat{F}_X(x_i), \hat{F}_Y(y_i)\}$ のプロット, (右) $\hat{\alpha}(u)$(赤) と $\alpha(u)$ の90%漸近信頼区間(黄).

大学共同利用機関法人 情報・システム研究機構
統計数理研究所
The Institute of Statistical Mathematics